DOCUMENT RESUME

ED 408 959 IR 018 372

AUTHOR	Washington, James M., Jr.
TITLE	Design Issues Adapting a Visual Paper-and-Pencil Test to the Computer: A Case Study--The Figure Classification Test.
PUB DATE	Jan 97
NOTE	9p.; In: VisionQuest: Journeys toward Visual Literacy. Selected Readings from the Annual Conference of the International Visual Literacy Association (28th, Cheyenne, Wyoming, October, 1996); see IR 018 353.
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Authoring Aids (Programming); *Computer Assisted Testing; Computer Interfaces; Computer System Design; Computer Uses in Education; Evaluation Methods; Instructional Materials; *Media Adaptation; Navigation (Information Systems); Optical Scanners; Pictorial Stimuli; Screen Design (Computers); Spatial Ability; *Test Construction
IDENTIFIERS	Digital Imagery; *Paper and Pencil Tests

ABSTRACT

This paper documents issues in converting the Figure Classification Test to the computer. The purpose of the test, which is almost entirely visual, is to determine the subject's ability to discover rules via the visual/spatial environment. The methodology of the paper-and-pencil Figure Classification Test is as follows: the subject views a series of two or three groups of pictures composed of printed shapes and is asked to classify each of the figures as belonging to one of three groups. In converting the paper-and-pencil test to the computer, the immediate concern was scanning the original drawings and converting them to line output. The philosophy of design for the project screens was to keep close to the paper version, yet to help the user navigate the test. The authoring aid used allowed mouse and keyboard input to operate equivalently for the user. Only minor changes to the original instructions were made; instructions to make the user familiar with the interface were added. Preliminary evaluation consisted of direct observation; 15 subjects at different times ran the program with no verbal instructions or coaching. The user interface and screen design seemed to be acceptable. No subject, after moving the mouse, ever attempted to use the keyboard for input. Users did not all gain immediate proficiency at the task, possibly due to the instructions. Device dependence issues also detracted from the computer test's equivalence to the original test. Direct observation of the program users appears to be a good first step toward improving and validating this computer-based test. (Contains 13 references.) (AEF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED 408 959

Design Issues Adapting a Visual Paper-and-Pencil Test to the Computer: A Case Study — The Figure Classification Test

By James M. Washington Jr.

Abstract

This paper documents some issues in a project assignment to convert the Figure Classification Test to the computer. The intent is to illuminate these issues, and to outline major questions. A brief description of the original paper-and-pencil test is followed by a description of the project's computer program, revealing part of the decision-making process that went into this implementation. Results from observation of initial users of the program are followed by some concluding thoughts. Although an initial philosophy of "faithfulness to the original test" produced an ultimately workable test project, observation of persons taking the computer-based test revealed opportunities for improvement.

Introduction

As effort in the field of teaching continues to move toward computerization, the idea of testing on the same platform follows predictably. Not only do we want to pre- and post-test the students to determine teaching effectiveness, but we may also be interested in some characteristics of the learners so that our methods may be better tailored to different learning styles.

The Paper-and-Pencil Test

The Figure Classification Test is a standard test published by the Educational Testing Service in Princeton, New Jersey (Copyright © 1962 by Educational Testing Service. All rights reserved.) The test's purpose is to determine the subject's *ability to discover rules that explain things*. The test is almost entirely visual, except for some initial textual instructions to the test subject and familiar control text like STOP. DO NOT GO ON TO THE NEXT PAGE UNTIL ASKED TO DO SO.

The methodology of the pencil and paper test is as follows:

The subject sees a series of two or three groups of pictures. Each group consists of three pictures, with each picture composed of printed shapes. The pictures are grouped according to some rule, such as "the shapes in Group 1 have gray shading, and, in Group 2, the shapes do not have gray shading," as illustrated in Figure 1. The subject is asked to classify each of the eight figures under the groups as belonging to one of the groups by writing a 1, 2, or 3 under the figure.

Figure 1 is admittedly a rather degenerate example. In the main body of the test, the test subject needs to make associations on visual and geometric concepts like right angles, parallelness, apparent dimensionality, relative spatial positioning, and so forth.

The Figure Classification Test was considered for computerization as part of a larger project. Abbot L. Packard, also a graduate student at Virginia Tech, was currently preparing a computer assisted tutorial / refresher on statistics for use by graduate students. As part of the project, he was interested in what learner characteristics make statistics more or less accessible. Since statistics may be considered a visual / spatial / rules course of study, a participant's score on the Figure Classification Test is interesting in this respect. Since the statistics tutorial is computer-based, it is desirable also to want to perform the learner testing on the

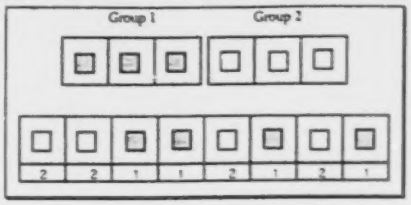
IR018372

BEST COPY AVAILABLE

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Alice D. Walker
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Figure 1

Example Figure Classification Test set Group 1 has gray shading. Group 2 does not.



same platform, hence the need for computerization of this test.

The Program

The task was conceptually easy: Take a standard, published and verified test, and put it on the computer. Display the test questions, get the user's input, and calculate the score, clearly merely a programming task. Display, user interface, and other design issues were "implementation details." The obvious strategy was to be as faithful as possible to the paper-and-pencil test, taking advantage of features and compensating limitations of the computer medium.

Initial Program Specifications

- Development environment: Authorware® Professional™ for Windows™
- The test: Copyright © 1962 by Educational Testing Service
- Philosophy: maintain faithfulness to original test
- Project completion: 3 weeks

Initial assumptions:

- Target resolution:
  VGA (640x480, 16 colors)
- User inputs:
  mouse / computer keyboard

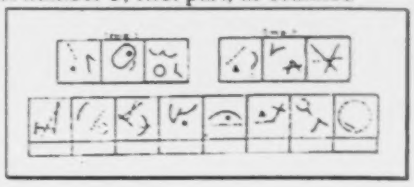
Getting the Graphics

Of immediate concern was the task of taking the original drawings and converting them for use in the project.

Scanning the test pages, then converting them to line output using a tracing package turned out to be unsatisfactory, as shown in Figures 2 and 3. Adding to this problem, some of the test's drawings use a dotted-line characteristic to distinguish a different "flavor" of line that sometimes is

Figure 2

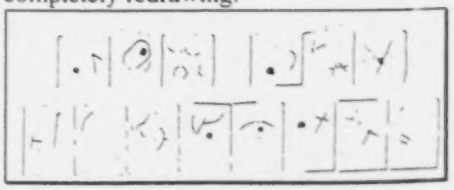
Set number 5, first part, as scanned



important to the rule for the set. Due to the relatively low resolution of the computer screen, dotted lines did not look good on the screen, and often broke up in ways that made the lines and curves much shorter. A small dotted-line circle often could not be identified as such due to this effect. Improving the resolution of the scan improved this somewhat, but also added extraneous photocopy artifacts that needed to be cleaned up using a bitmap editor.

Figure 3

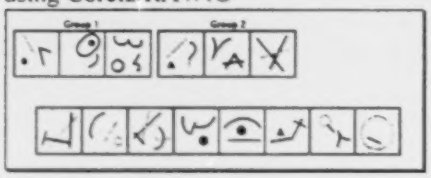
Set number 5, first part, as "traced." Cleaning up this mess took longer than completely redrawing.



Ultimately, completely redrawing the test became the easiest option, in terms of speed and quality. Dotted lines and curves became gray lines and curves. The "Layers" feature of CorelDRAW!® allowed use of a standard template for the boxes. The individual drawings were redrawn within these, as shown in Figure 4. From there, the figures could be copied-and-pasted into Authorware®.

Figure 4

Set number 5, first part as redrawn using CorelDRAW!®



The original test centers the group figures above the eight test figures. For consistency of user interface for the project, Groups 1 and 2 always remain in the same place relative to the test figures, regardless of whether a third group is present.

Since the redrawing was performed by persons knowledgeable about the rules for the particular sets, conversion of dotted lines to gray and the standard placement of the groups are the only substantive visual differences between sets in the original paper-and-pencil test and the sets in the project version.

Screen Design

The philosophy of design for the project screens was to keep close to the paper version, yet to help the user navigate the test. In the paper test, up to four test sets are displayed per page. For the project, only one set is displayed per screen, in deference to the relatively lower resolution of the computer screen.

As designed, the top of the screen displays status information, the center displays the active zone for user input, and the bottom of the screen contains navigation buttons. Feedback and instructions to the user also are important, so interactive prompting and highlighting of user selections occurs while the test is being performed. Figure 5 shows an example screen, as the user sees it before selection of a test figure, and Figure 6 shows the same screen after user selection of a figure.

Authorware® provides easy access to user interface methods, so mouse and keyboard inputs operate equivalently for the user. To use the mouse, the user clicks anywhere within the test figure, then anywhere within the group to which the test figure belongs. To use the keyboard, the user presses the number on the keyboard associated with the test figure, then the number of the group. To move on, the mouse user clicks on the "Go On to Next Set" button, and the keyboard user presses "Enter."

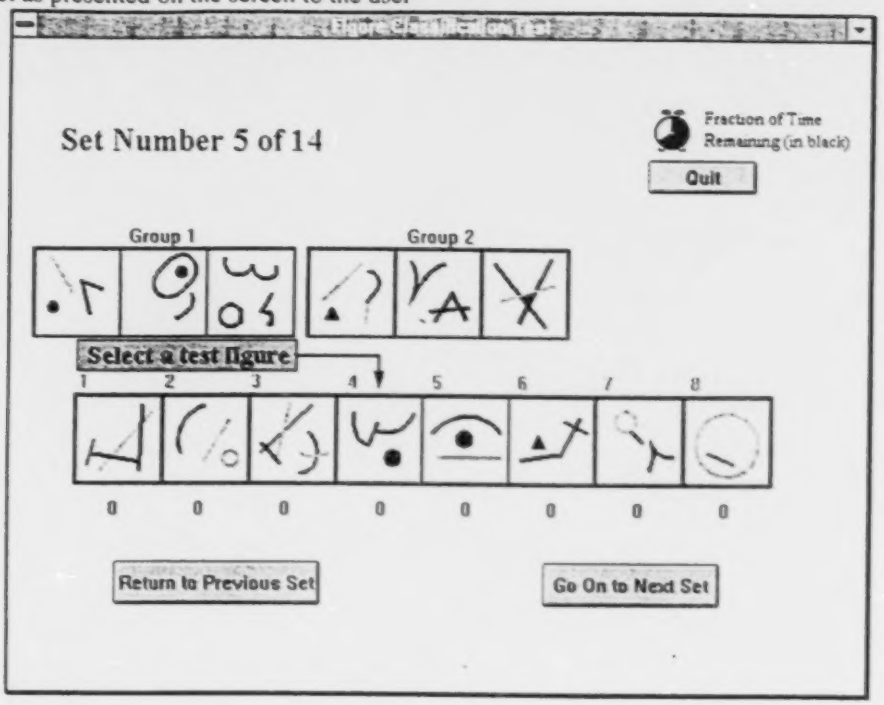
Since the original test allows the subject to return to previous sets within the current test part, the project test also allows for this with a button for "Return to Previous Set."

The original test is timed, eight minutes for each of the two parts of the test. This is also handled easily by Authorware®. The clock device in the upper right corner of the screen is provided by the authoring package, though placement and usage on the screen are left for the programmer to decide.

BEST COPY AVAILABLE

Figure 5

A set as presented on the screen to the user


Set Number 5 of 14
Fraction of Time Remaining (in black)
Quit
Group 1
Group 2
Select a test figure
1 2 3 4 5 6 7 8
0 0 0 0 0 0 0 0
Return to Previous Set
Go On to Next Set

Instructions

In keeping with the philosophy of staying as close as possible to the written test, only minor changes to the instructions were made. These changes served only to make the user familiar with the interface, assuming some familiarity with computer terms like "click with the mouse" and "press the key." The original sample problems are presented in the same format as the test proper. The first sample is completed for the user, and the second sample has a "Hint" button that, when pressed, generates a display of the correct answers and the explanation of the rules for the set.

Preliminary Evaluation

Preliminary evaluation consisted of direct observation of persons using the program. Fifteen subjects at different times sat in front of a computer to run the program with no verbal instructions or coaching, while being observed by the programmer. Subjects ranged in age from 20 to 40 years old, and rated their computer familiarity in the range of "familiar" to "expert." No statistics have been generated, but these observations have led to some questions that indicate a need for further study and improvement of the current design.

User Interface

The user interface and screen design seem to be acceptable. Clicking a figure, then clicking a group seemed to work well for all subjects. The subjects saw a "video game" metaphor, which was familiar to the extent that clicking with the mouse in a region generally gives a predictable response.

The current design makes mouse usage mandatory and keyboard input optional. A mouse click is mandatory on the buttons that say "Go On to Next Part," a

design feature that prevents those who use the keyboard from blindly pressing "Enter" after the last set of a part, removing their opportunity to review or complete previous responses. Mouse use is also mandatory for the button to "Return to Previous Set," for no apparent reason, however, no user has yet asked for a keystroke to do this.

It may be a general rule that if mouse use is mandatory in any respect, users will not tend to use the keyboard. No subject, after moving the mouse, ever attempted to use the keyboard for input. When asked why they used the mouse in preference to the keyboard, most subjects said it was just preferred. It is noted that more mental effort is needed to code the test figure locations to the numbers 1 through 8 and the group locations to the numbers 1 through 3 than moving the mouse cursor to those locations.

Users did not all gain immediate proficiency at the task, however. The instructions provided in the test may be insufficient. A few subjects, even after the second sample problem, still stated that they did not understand what they were supposed to do. All of the subjects who were ignored on this issue (and it was difficult not to answer questions) eventually did figure out the task. It appears, particularly with computer-based testing, where no assumption can be made about the availability of a person to answer questions, that instructions must be abundantly clear. This is not just a different test; it is a new computer program that must be learned.

Placement of the "Quit" button was problematic. Originally, it was placed with the navigation keys at the bottom of the screen, but users often pressed it instead of the "Go On to Next Set" button. These users reported that they knew what the button was before they pressed it, but pressed it anyway. These

Figure 6

Screen after selection of test.
The highlight around test selection 6 is red on the screen

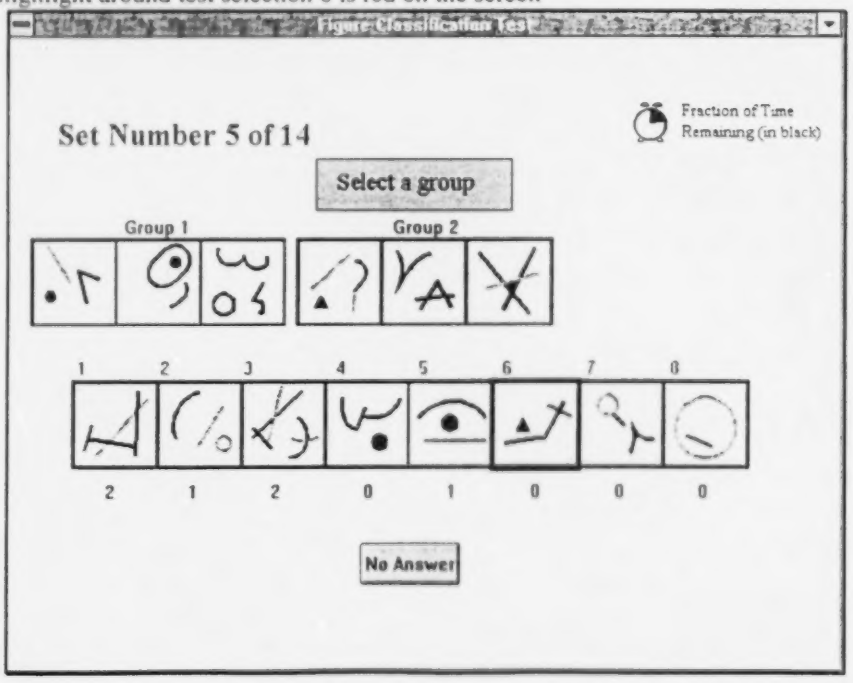


Figure Classification Test
Set Number 5 of 14
Fraction of Time Remaining (in black)
Select a group
Group 1
Group 2
1 2 3 4 5 6 7 8
2 1 2 0 1 0 0 0
No Answer

BEST COPY AVAILABLE

users had just completed using another package with the "Go On" button in the same place as the "Quit" button in this application. Navigation and exit keys should be spaced apart from one another, and should, if possible, be consistent between applications.

Device Dependence

For a program developed in Authorware®, the project is surprisingly device-dependent. Of particular concern is the difference in contrast between monitors. Gray shading can turn black or disappear completely depending on the position of a knob on the user's monitor. Since the difference between black, white, and gray is essential to the test, would another color be better than the gray?

Another device-dependence question is the issue of screen curvature. Is there a limit on the curvature of a computer monitor at which parallel lines and right angles no longer appear to be parallel lines and right angles?

Still another portability question is the speed of the computer with respect to screen updates. The project uses the times allowed from the original test, and does not allow extra time for updating the screen between sets. Moving from one screen to another on a 486/33 machine takes approximately one second. Screen updates using a faster or slower machine with a different video card would necessarily take a different amount of time. Is this important to the validity of the test?

Test Factors

Some users appeared to be more proficient than others at using the mouse and keyboard. Is this important to the user's score on the test?

The test is presented linearly, with one set following another set in the original order. Is this ordering important? How was the original order determined? Some sets do appear to be more difficult than others, since the users seem to spend more time on some sets than others. Does the computer screen presentation change factors that affected the original ordering?

Most users worked the test linearly and were hesitant to move on to the next set if they had trouble. Thus, it appears that the project also tests a subject's willingness to give up and move on. Is this effect enhanced by this particular presentation mode? Was this part of the original design? Would clearer instructions help? Would presentation of a "difficulty index" help?

The project provides instant scoring to the user. After the test, most users asked what the score meant. Since this is a test that can be taken only once due to learning factors, would it be appropriate to give the user some idea of the significance of the user's score compared with others? How significant is the score? Should high school students be grouped with college students? Should engineering / technical persons be grouped with education / social sciences persons?

Testing Conclusions

The project presents a substantially different test than the original paper-and-pencil test. It tests not only the subject's *ability to discover rules that explain things*, but other factors, like willingness to learn a new computer program, mouse / keyboard skills, and propensity to give up and move on. Device dependence issues also detract from the computer test's equivalence to the original test. To its credit, however, it presents a substantially level playing field to the user. Persons with more difficulty *discovering rules that explain things* spent more time on the initial sets and

BEST COPY AVAILABLE

had lower scores than persons with less difficulty. None of the observed subjects completed all sets, indicating that the current time limit is a limiting factor to the higher scores. It seems likely that the current configuration of the project will significantly differentiate between persons with high ability and low ability in the test's design purpose, although the scores will not likely be comparable to the original test's scores.

Conclusions

There is more than meets the eye in adapting a visual test to the computer. An attempt at faithfulness to the original paper-and-pencil test may not be sufficient to make the subject's score the same from paper to computer. The author's experience in this regard is not unique. Review of publications on the subject of computer-based testing indicates that even text-based psychological tests on the computer may show different results than the paper-and-pencil versions. Themes in these papers include computer anxiety (can this be distinguished from anxiety in learning a new computer program?), the primacy of the importance of the user interface, computer response time, programming errors, and miscommunication between the test preparer and the programmer.

It is perhaps an overstatement that computer-based tests are a new frontier in testing. Computer-based tests have been around since the advent of computers. Recall that almost everyone's second programming task involves quizzing the user for his name. Research into quality and design issues related to computer-based testing, however, has only taken place in the last fifteen years or so.

Direct observation of the users of the program appears to be a good first step toward improving and validating this computer-based test. More work will be necessary to make the project comparable to the original test, if that is a desirable characteristic for the project. Among the possible strategies are:

- improved instructions, perhaps with more examples
- a better representation of "flavor" than gray
- timing the responses to determine a better order for the test
- randomizing the sets (would this improve or hurt scoring?)
- establishing a time limit for comparable scores with the paper test
- replacing or redrawing sets that prove to be more difficult to the users
- shortening or breaking up the test
- accumulation of scores for persons of different backgrounds for comparison and assignment of more meaningfulness to the user's score

It is possible that a complete redesign of the test might be a desirable option. The paper-and-pencil test is limited to black ink on white paper and geometrical shapes; would additional colors or motion or more iconic clues enhance or detract from the test purpose? Could such a test be adapted to give additional information about the subject?

As the title suggests, this paper is about issues, not answers. The computer is a promising platform for testers and test-takers alike. For the user, tests can become more like video games, and the public seems to have an almost inexhaustible appetite for these, which are, after all, tests of some sort. For the tester, the computer also provides the ability to diagnose and improve a test to a degree never before possible. It is only a

matter of doing so. Only after the limitations of a test are known can they be determined to be important, unimportant, worth pursuing, or worth a complete rethink of the project.

Bibliography

Beaumont, J. G., (1981) Microcomputer-aided assessment using standard psychometric procedures. Behavior Research Methods and Instrumentation; 1981 13

Booth, J. (1991) The key to valid computer-based testing: The user interface. Special issue: Computerized diagnostics. European Review of Applied Psychology; 1991 Vol 41(4) 281-283

Dimock, P. H. and Cormier, P.(1991) The effects of format differences and computer experience on performance and anxiety on a computer-administered test. Measurement and Evaluation in Counseling and Development; 1991 Oct Vol 24(3) 118-126

George, C. E., Lankford, J. S., and Wilson, S. E. (1992) The effects of computerized versus paper-and-pencil administration on measures of negative affect. Computers in Human Behavior; 1992 Vol 8(2-3) 203-208

Johnson, J and Johnson, K. (1981) Psychological considerations related to the development of computerized testing stations. Behavior Research Methods and Instrumentation; 1981 13

Kaplan, H (1992) Representation of on-line questionnaires in an editable, auditable database. Behavior Research Methods and Instrumentation; 1992, 24(2) 373-384

Knupfer, N. (1994) Computers and visual learning. Visual Literacy: a spectrum of visual learning. D. M. Moore and F. M. Dwyer, editors (1994) Educational Technology Publications, Inc., Englewood Cliffs, NJ

Kyllonen, P. C. (1991) Principles for creating a computerized test battery. Intelligence; 1991 Jan-Mar Vol 15(1) 1-15

Labre, M., Clements, N., Fitzhugh, K., Lancelotta, G., et al. (1987) The effect of computer administered testing on test anxiety and performance. Journal of Educational Computing Research; 1987 Vol 3(4) 428-433

Most, R. (1987) Levels of error in computerised psychological inventories. Special Issue: Computerised psychological testing. Applied Psychology An International Review; 1987 Vol 36(3-4) 376-383

Nurius, P. S. (1990) Computer literacy in automated assessment: Challenges and future directions. Computers in Human Services; 1990 Vol 6(4) 283-287

Rafaeli, S., and Tractinsky, N. (1989) Computerized tests and time: Measuring, limiting, and providing visual cues for response time in on-line questioning. Behaviour and Information Technology; 1989 Sep-Oct Vol 8(5) 335-351

Schoenfeldt, L. (1989) Guidelines for computer-based psychological tests and interpretations. Computers in Human Behavior; 1989 Vol 5(1) 13-21

BEST COPY AVAILABLE